

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:37 ; Search time 170.72 Seconds

27.443 Million cell updates/sec

Title: US-09-331-631A-5\_COPY\_76\_144

Sequence: 1 NRQRDPQQYECCCKRCRR...EEQREDEKYEERMKEGDN 69

Scoring table: BLOSUM62

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Maximum Match 1008

```
Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*
```

**Pred. No.** Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	173	45.4	566	2	S22477	vicilin precursor
2	147.5	38.7	509	2	S08059	alpha-globulin type
3	145.5	38.2	588	1	FWCNAB	alpha-globulin b p
4	128	33.6	605	2	S06398	alpha-globulin t y p
5	115	30.2	810	2	T44430	protein pvi100 (imp
6	114.5	30.1	1038	2	T02634	rep protein homoid
7	105.5	27.7	385	2	T11923	hypothetical prote
8	103	27.0	1898	1	A45973	trichovallin - hum
9	102	26.8	613	2	S27770	trichovallin - hum
10	101	26.5	1737	2	A59335	hypothetical prote
11	100	26.2	1407	1	S28589	unconventional myo
12	99	26.0	1390	2	T14004	trichovallin - rab
13	97.5	25.6	839	2	I50590	trifa protein - sili
14	97.5	25.6	877	2	I50591	Class I INCENP pro
15	97.5	25.6	1027	2	T46481	Class II INCENP pr
16	97.5	25.6	1233	2	T30989	hypothetical prote
17	95.5	25.1	483	2	T06459	serine/threonine p
18	95.5	25.1	1023	2	S12519	62k sucrose-bindin
19	95	24.9	524	2	J01730	glutacitin - fruit
20	95	24.9	905	1	RG8Y55	62k sucrose-bindin
21	94	24.7	233	2	T17218	regulatory protein
22	94	24.7	648	1	J01150	hypothetical prote
23	94	24.7	1085	2	S62516	protein kinase (EC
24	93.5	24.5	781	2	T02272	hypothetical colla
25	93	24.4	139	2	A26892	hypothetical prote
26	93	24.4	429	2	S29565	Mopa box protein -
27	93	24.4	533	2	T24961	apolipoprotein A-I
28	93	24.3	849	2	S61962	hypothetical prote
29	92.5	24.3	1403	2	S24548	probable membrane
						homoclit protein p

30	91.5	24.0	406	2	T24492	hypothetical prote
31	91.5	24.0	540	2	S21825	viciolin-like stora
32	91.5	24.0	573	2	A53234	globulin-1S, GLB1S
33	91	23.9	385	2	T20410	hypothetical prote
34	90	23.6	678	2	A54514	glutamic acid-rich
35	90	23.6	1346	2	T42637	hypothetical prote
36	89.5	23.5	758	2	S34522	hypothetical prote
37	89	23.4	285	1	I46207	involucrin - dog
38	89	23.4	582	2	B53234	viciolin-like stora
39	88.5	23.2	407	2	T03258	globulin - maize
40	88	23.1	292	2	J60233	tropoin-1 - scall
41	88	23.1	447	2	S52391	centrosomin B - mo
42	88	23.1	766	2	G54024	protein kinase (EC
43	88	23.1	768	2	H54024	protein kinase (EC
44	88	23.1	777	2	F54024	protein kinase (EC
45	88	23.1	779	2	E54024	protein kinase (EC

## ALIGNMENTS

```

RESULT 1
S22477
vicillin precursor - cacao
C:Species: Theobroma cacao (cacao)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: S22477; S22478; S18105; S2050
R:McHenry, L.; Fritzt, P.J.
Plant Mol. Biol. 18, 1173-1176, 1992
A:Title: Comparison of the structure and nucleotide sequence of vicillin genes of cacao
A:Reference number: S22477; MUID:92286309
A:Accession: S22477
A:Molecule type: DNA
A:Residues: 1-566 <MCH>
A:Cross-references: EMBL:X62625
A:Accession: S22478
A:Molecule type: mRNA
A:Residues: 1-452 <MC2>
A:Cross-references: EMBL:X62626
C:Genetics:
A:introns: 211/1; 269/3; 296/3; 391/3; 502/1
C:Superfamily: glycinin
C:Keywords: seed; storage protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-566/Product: vicillin #status predicted <MAT>

Query Match 45.4%; Score 173; DB 2; Length 566;
Best Local Similarity 32.6%; Pred. No. 1.5e-07;
Matches 31; Conservative 17; Mismatches 15; Indels 32; Gaps 1;

OY 3 QRDPOOYECCORRCORETEPRHMQICORCERRYEKKRKO----- 46
:|||||:|||||: || : :|||||:|:|:|:
Db 35 ERDPQOYECCORRCESATEEREEOECORCERPEYKQROQEBELQROYQOCGRCOE 94
OY 47 -----KRYECCORPEDEKYEERMK 65
::|:|:|:|:|:|:|

Db 95 QOQCGREDOCCORCKCWQYKFEORGEHENVHNNKK 129

RESULT 2
S08059
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
N:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S08059
R:Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.
A:Reference number: S08398
A:Accession: S08059
A:Status: not compared with conceptual translation

```



Query Match 30.1%; Score 114.5; DB 2; Length 1038;

Best Local Similarity 36.9%; Pred. No. 0.017;

Matches 24; Conservative 23; Mismatches 17; Indels 1; Gaps 1;

QY 3 QRDPOQOYEDCCRCORRETEPRHMQICQRCRRRYEKRRKQKRYEEO-QREDEEKE 61

DB 936 QOEQOEQOEQERREDEQOEQOEQOEQOEQOEQOEQOEQOEQOEQOEQOEQOEQOE 995

QY 62 ERKKE 66

DB 996 QREDE 1000

RESULT 7

T19201 hypothetical protein C11G6.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T19201

R:Milkinson, J.

Submitted to the EMBL Data Library, March 1996

A:Reference number: Z19089

A:Accession: T19201

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-385 <MIL>

A:Cross-references: EMBL:Z70204; PIDN:CAA94113.1; GSPDB:GN00028; CESP:C11G6.3

A:Experimental source: clone C11G6

A:Gene: CESP:C11G6.3

A:Genetics:

A:Map position: X

A:Introns: 12/1; 106/3; 171/2; 335/2

Query Match 27.7%; Score 105.5; DB 2; Length 385;

Best Local Similarity 28.4%; Pred. No. 0.042;

Matches 21; Conservative 24; Mismatches 20; Indels 9; Gaps 1;

QY 2 RORDPOQOYEDCCRCORRETEPRHMQ-----ICQRCRRRYEKRRKQKRYEEO 52

DB 146 RERDERERERQOEKREDAAREIEKAKMDAKRYAEEERKRKRREKKQKE 205

QY 53 QREDEEKEERKKE 66

DB 206 LKEKERSEKKE 219

RESULT 8

A45973 trichohyalin - human

C:Species: Homo sapiens (man)

C:Date: 03-May-1994 #sequence\_revision 01-Mar-1996 #text\_change 22-Jun-1999

C:Accession: A45973

R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.

J. Biol. Chem. 268, 12164-12176, 1993

A:Title: The structure of human trichohyalin. Potential multiple roles as a functional E

ed (cross-linking) protein.

A:Reference number: A45973; MUID:93280194

A:Accession: A45973

A:Molecule type: DNA

A:Residues: 1-1898 <LEE>

A:Cross-references: GB:L09190; NID:9292835; PIDN:AA65582.1; PID:9292836

C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh

Covalent modifications to this protein include conversion of arginine to citrulline and

C:Genetics:

A:Gene: GDB:THH

A:Cross-references: GDB:136223; OMIM:190370

A:Map position: 1q21-1q21

C:Superfamily: trichohyalin; calmodulin repeat homology

C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 27.0%; Score 103; DB 1; Length 1898;

Best Local Similarity 31.0%; Pred. No. 0.25;

Matches 22; Conservative 23; Mismatches 20; Indels 6; Gaps 1;

QY 2 RORDPOQOYEDCCRCORRETEPRHMQICQRCRRRYEKRRKQKRYEEOQRE 55

DB 267 RQRELDQEEELRLKRLQGLRRLRRQEEQOQRLRLRRQRLRRQEEERRRQOERRRQGE 326

QY 56 DEEKEERKKE 66

DB 327 RREQOEEERREQ 337

RESULT 9

S27770 hypothetical protein 1 - African malaria mosquito (fragment)

C:Species: Anopheles gambiae (African malaria mosquito)

C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Sep-1997

C:Accession: S27770

R:Basansky, N.J.; Paskewitz, S.M.; Mills-Hamm, D.M.; Collins, F.H.

Submitted to the EMBL Data Library, June 1992

A:Description: Distinct families of site-specific retroposons occupy identical positi

A:Reference number: S27770

A:Accession: S27770

A:Molecule type: DNA

A:Residues: 1-613 <BAS>

A:Cross-references: EMBL:M93690; NID:g159615; PID:g159616

Query Match 26.8%; Score 102; DB 2; Length 613;

Best Local Similarity 26.5%; Pred. No. 0.12;

Matches 18; Conservative 27; Mismatches 23; Indels 0; Gaps 0;

QY 2 RORDPOQOYEDCCRCORRETEPRHMQICQRCRRRYEKRRKQKRYEEOQREDEEKE 61

DB 252 QQRNQOEQEMQOQOQOQHQHQREQOQOQRVQOQOQHQHQOQOQOQOQOQOQOQOQOQOQELMT 311

QY 62 ERKKEGDN 69

DB 312 TVVRRRON 319

RESULT 10

A59235

unconventional myosin heavy chain MYOM - slime mold (Dictyostelium discoideum)

C:Species: Dictyostelium discoideum

C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 08-Sep-2000

C:Accession: A59235

R:Geisler, H.; Schwarz, E.C.; Soldati, T.

Submitted to Genbank, September 1998

A:Description: Identification of two novel and highly divergent myosins in Dictyostel

A:Reference number: A59235

A:Accession: A59235

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1737 <GET>

A:Cross-references: GB:AFO90533; NID:g5714395; PIDN:AA47903.1; PID:g5714396

A:Experimental source: strain AX2

C:Genetics:

A:Gene: myom

A:Map position: 6, aldb-cab2

C:Superfamily: myosin motor domain homology

F:62-874/Domain: myosin motor domain homology #status atypical <MNO>

Query Match 26.5%; Score 101; DB 2; Length 1737;

Best Local Similarity 35.8%; Pred. No. 0.34;

Matches 24; Conservative 20; Mismatches 19; Indels 4; Gaps 2;

QY 1 NRORDPOQOYEDCCRCORRETEPRHMQICQRCRRRYEKRRKQKRYEEOQREDEEKE 59



submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23035  
A:Accession: T46481  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1027 <AAAs>  
A:Cross-references: EMBL:AL137755  
A:Experimental source: adult testis; clone DKFZp434A025  
C:Genetics:  
A:Note: DKFZp434A025.1

Query Match 25.6%; Score 97.5; DB 2; Length 1027;  
Best Local Similarity 39.2%; Pred. No. 0.43;  
Matches 29; Conservative 17; Mismatches 17; Indels 11; Gaps 4;

OY 2 RQDPDOOYECCOK--RCQRETEPRHMQIQRCERRYEKEKRR---QOKRYEEOOR- 54  
||:|||||:: ||||| | | | | |::|||: ::| ||::|  
Db 134 RQRIEQQKQRRRLQEQRREREARQQQERQOR--RRQEEKRRLLELERRRKEEERR 191

OY 55 --EDDEKYEEERKE 66

Db 192 RAEEKRRVEREQE 205  
|::: ||:|

Search completed: March 1, 2001, 15:52:38  
Job time: 562 sec

